

Описание задачи

В сервисе «Яндекс Учебник» требуется предсказать **отток учителей** на ближайших горизонтах. По истории активности на платформе и агрегированным поведенческим признакам нужно оценить вероятность, что учитель прекратит активное использование сервиса:

- `week_churn` — отток в течение **недели**,
- `month_churn` — отток в течение **месяца**.

Задача формулируется как **бинарная классификация** по табличным числовым признакам и эмбедингам (два независимых таргета).

Данные

Все события и тексты уже преобразованы в числовые признаки. Для каждого учителя и даты среза доступны:

- `nid` — идентификатор учителя;
- `report_date` — дата формирования среза;
- `v_0 ... v_N` — числовые признаки (агрегаты активности, эмбединги поведения и т.п.);
- `week_churn`, `month_churn` — целевые метки (0 — не ушёл, 1 — ушёл) — **только в `train.csv`**.

Данные анонимизированы.

Формат файлов

- **`train.csv`** — обучающая выборка: `nid`, `report_date`, набор фич `v_*`, таргеты `week_churn`, `month_churn`.
- **`test.csv`** — тестовая выборка: те же признаки (`nid`, `report_date`, `v_*`), **без** таргетов.
- **`sample_submission.csv`** — пример файла для отправки: `nid,week_churn,month_churn` со значениями-заглушками.

Разделитель во всех файлах — запятая.

Цель и метрика

Нужно предсказать **вероятности** оттока для объектов из `test.csv` по обоим целям.

Метрика качества — ROC-AUC (macro):

$$\text{score} = \frac{1}{2} (\text{ROC_AUC}(\text{week_churn}) + \text{ROC_AUC}(\text{month_churn}))$$

Формат отправки

Файл `submission.csv` со следующими колонками **в строгом порядке**:

`nid, week_churn, month_churn`

Требования:

- длина и порядок строк **в точности совпадают** с `test.csv` (оценивание построчное);
 - значения `week_churn` и `month_churn` — **вещественные вероятности** в диапазоне **[0, 1]**;
 - отсутствуют NaN и числа вне диапазона;
 - дополнительные колонки не допускаются;
 - повторяющиеся `nid` допускаются, если они присутствуют в `test.csv`.
-

Правила

- Разрешены любые стандартные ML-библиотеки и модели (линейные, деревья, бустинги, нейросети и т.д.).
 - Внешние данные и дополнительная ручная разметка **запрещены**.
 - Решение должно быть **воспроизводимо** (фиксируйте `random_state`, версии библиотек).
-

Оценивание

Участники ранжируются по **убыванию ROC-AUC (macro)** на скрытой тестовой выборке.